REMARKS

Applicants acknowledges receipt of the Final Office Action dated March 18, 2003. In that action, the Examiner: 1) objected to claims 16, 17 and 19-20 for various informalities; 2) rejected all the claims as allegedly being directed to non-statutory subject matter; 3) rejected claim 16 as allegedly anticipated by *Dhillon* (U.S. Patent No. 6,269,376); 4) rejected claims 1-15 as allegedly unpatentable over *Guha* (U.S. Patent No. 6,092,072); 5) rejected claims 17-20 as allegedly unpatentable over *Dhillon* in view of *Guha*; and 6) made the action Final.

With this Preliminary Amendment, Applicants amend claims 1-10 and cancel claims 11-20. Applicants believe the pending claims are allowable over the art of record and respectfully request reconsideration.

I. CLAIM REJECTIONS

A. Claim 1

Claim 1, as amended, is directed to a computer readable medium containing a program executable by a microprocessor. When executed, the program performs a method for clustering data comprising receiving a plurality of data points for clustering, receiving a size parameter for specifying the number of data points to be moved at one time, clustering the data points by using the size parameter to generate clustered results, and determining whether the clustered results are satisfactory. When the clustered results are satisfactory, the clustering stops, and when the clustering results are not satisfactory, the size parameter is revised and clustering again performed based on the revised size parameter. The Examiner rejected claim 1 as allegedly directed to non-statutory subject matter, and as allegedly anticipated by *Guha*.

Applicants have amended claim 1 to more clearly define that the method steps recited therein are program executable by a microprocessor. Applicants respectfully submit that one of ordinary skill in the art, after reading the specification, would clearly understand the methods described therein to be directed for use in a computer system. In this regard, Applicants' specification included the following statements:

> There are numerous applications that can utilize the aggregate clustering method and system of the present invention to cluster data. For example, these applications include, but are not limited

> to, **data mining applications**, customer segmentation applications, document categorization applications, scientific data analysis applications, **data compression applications**, vector quantization applications, and **image processing applications**.

Specification, page 16, lines 9-13 (emphasis added). Although all these examples would be read by one of ordinary skill in the art to be applications executed on a computer system, the data mining applications, data compression applications, and image processing applications are clearly directed to use with computer systems. Thus, Applicants respectfully submit that no new matter is presented by the amendments. Applicants make these amendments only to address the Examiner's non-statutory subject matter rejection, and not to define over the *Guha* reference.

Further, Applicants respectfully submit that one of ordinary skill in the art would be enabled by the specification. Figures 2 and 3 of the originally presented specification are flow charts delineating the steps to implement in performing the method. Flow charts are, of course, a preliminary step in coding software to perform a function. Thus, Applicants submit one of ordinary skill could, utilizing the flow charts and the detailed explanation of equations to perform the clustering (starting page 14 of the specification) code a software program to perform the claimed steps.

Applicants respectfully submit that the *Guha* reference fails to teach or render obvious the limitations of claim 1. In the Office Action dated March 18, 2003, the following statements are made:

> [I]n response to applicant's argument that the reference fails regarding the size parameter of the invention, examiner holds that Guha's C parameter is essentially the same as the claimed size parameter. The C parameter specifies the number of data points that will be evaluated when deciding whether to merge a pair of clusters .... The merge procedure of Guha is essentially the same as the claimed move.

Serial No.: 09/684,488
Filed: October 4,2000

Office Action dated March 18, 2003, pages 2-3. Applicants respectfully disagree that this is the teaching of *Guha*. At most, the *Guha* reference may teach that only pairs of clusters should be merged.[1]

> [E]ach successive step merges the **closest pair** of clusters ... .

Guha, Col. 6, lines 50-51 (emphasis added).

> [S]tarting with the individual points as individual clusters, at each step the **closest pair** of clusters is merged to form the new cluster. The process is repeated until there are only k remaining clusters.

Guha, Col. 7, lines 18-21 (emphasis added).

> At step 209, the merge procedure ... is used to merge the **closest pair** of cluster u and v ... .

Guha, Col. 8, lines 17-19 (emphasis added). At best, *Guha* may teach only merging **pairs of clusters**.

With respect to the c parameter of *Guha*, this parameter appears to only be used in determining a set of representative points. The *Guha* reference may use only two data structures, a heap and a k-d tree.

> The method of the present invention makes extensive use of two data structures, a heap and a k-d tree. ... [C]orresponding to every cluster u, there exists a single entry in the heap; the entries for the various clusters u are arranged in the heap in the increasing order of the distances between u and u.closest.
>
> The second data structure is a k-d tree that stores the **representative points** for every cluster. ... When a pair of clusters is merged, the k-d tree is used to compute the closest cluster for clusters that may previously have had one of the merge clusters as its closest cluster.

Guha, Col. 7, lines 41-59 (emphasis added). Thus, the k-d tree stores representative data points that are used for calculating distances. The question becomes then, how are these representative points calculated?

> In order to compute the distance between a pair of clusters, for each cluster, **representative points** are stored. These [representative points] are determined by first choosing a constant number c of well scattered points within the cluster, and then

---

[1]   In *Guha*, each individual point of the data is considered to be a "cluster," and once combined, the combined element is likewise called a "cluster." *Guha*, Col. 6, lines 49-52.

> shrinking them toward the mean of the cluster by a fraction α. ...
> Thus, only the c representative points of a cluster are used to
> compute the distance from the other clusters.

Guha, Col. 6, lines 52-61 (emphasis added).

> At step 340, the procedure has created a set of **representative
> points** w.rep for the new cluster w **that were initially selected as c
> well-scattered points** ... .

Guha, Col. 10, lines 26-30 (emphasis added).

Thus, the parameter c in the *Guha* reference is not used to determine how
many points should be moved from cluster to cluster; rather, the parameter c is
used only to determine **how many of the points should be used to calculate a
representative point for distance calculation** purposes. Thus, *Guha* does not
teach that, "the C parameter specifies the number of data points that will be
evaluated **when deciding to merge a pair of clusters** ... ." Further, *Guha* does
not teach "the size parameter (C) specifies the number of data points to be
moved at one time from one cluster to another cluster ... ."

Thus, Applicants respectfully submit that *Guha* does not teach, suggest, or
even imply a size parameter for specifying the number of data points to be moved
at one time. The c parameter of *Guha* is only used in calculating representative
points. *Guha* expressly discloses combining pairs of clusters, and the c
parameter does not change this teaching.

Based on the foregoing, Applicants respectfully submit that claim 1, and all
claims which depend from claim 1 (claims 2-10), should be allowed. In a fashion
similar to the amendments to claim 1, claims 2-10 were amended to more clearly
indicate that the method steps described therein are for use in a computer
system.

B.   Claim 7

Claim 7 is directed to a computer readable medium having all the
limitations of claim 1 and further requiring decreasing the size parameter. Claim
7 was rejected as allegedly obvious over *Guha*. Applicants amended claim 7 to
more clearly define that the methods may be used within a computer system, and
not to define over the *Guha* reference.

Serial No.:   09/684,488
Filed: October 4, 2000

As discussed with respect to claim 1, the c parameter of *Guha* is used to calculate a representative point of data points within a cluster, and not as an indication of how many data points should be put in a cluster. With regard to a value of the c parameter in the *Guha* reference, *Guha* makes the following statement, "Preferably, c will have a value greater than 10." *Guha*, Col. 10, line 11. By contrast, claim 7 specifically requires that the size parameter is decreased to make fine adjustments to the clustering. *Guha* does not teach or render obvious all the limitations of claim 7 if the smallest granularity that may be used is 10 data points.

In forming the rejection of claim 7, the *Guha* reference is characterized as teaching that "c is shrunk toward the mean by a fraction α ... ." Office Action dated September 11, 2002, page 7. Applicants respectfully disagree. It is not the c parameter of *Guha* that is shrunk by the fraction α, it is the chosen scattered points. In fact, the locations cited makes this exact statement:

> The chosen scattered points are next shrunk towards the mean of the cluster by a fraction α.

Guha, Col. 4, lines 36-38.

> [A] constant number c of well scattered points that capture the shape and extent of the cluster.

Guha, Col. 4, lines 35-36. Thus, the teachings of *Guha* are deficient.

Claim 7 is dependent from claim 1 and is allowable for at least the same reasons, as well as the additional limitation regarding decreasing the size parameter, which the Examiner expressly states is not taught by the *Guha* reference.

II.   CLAIM CANCELLATIONS

With this Office Action Response, Applicants have cancelled claims 11-20. Applicants make this cancellation to narrow the issues before the Examiner. This cancellation is not without prejudice to later asserting these claims here, in a divisional application, a continuation-in-part application, or the like.

III.   EXAMINER INTERVIEW

On Wednesday, March 23, 2003 the undersigned conducted a telephonic interview with Examiner Hamilton.

The Examiner was provided a set of discussion materials prior to the interview that included some of the claims of the present application with proposed amendments. The discussion material also comprised selected sections and arguments regarding one of the references used in the Final Office Action.

Claims 1-10 were discussed in general.

The *Guha* reference was discussed.

The general thrust of the principle arguments were directed to the patentability of the claims in light of the specification.

In the interview, the undersigned and the Examiner also discussed whether the Office Action dated March 18, 2003 was properly final.

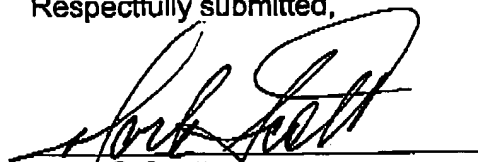No agreements were reached.

IV. CONCLUSION

Applicants respectfully request reconsideration and allowance of the pending claims. If the Examiner feels that a telephone conference would expedite the resolution of this case, he is respectfully requested to contact the undersigned.

In the course of the foregoing discussions, Applicants may have at times referred to claim limitations in shorthand fashion, or may have focused on a particular claim element. This discussion should not be interpreted to mean that the other limitations can be ignored or dismissed. The claims must be viewed as a whole, and each limitation of the claims must be considered when determining the patentability of the claims. Moreover, it should be understood that there may be other distinctions between the claims and the prior art which have yet to be raised, but which may be raised in the future.

Serial No.: 09/684,488
Filed: October 4, 2000

If any fees or time extensions are inadvertently omitted or if any fees have been overpaid, please appropriately charge or credit those fees to Hewlett-Packard Company Deposit Account Number 08-2025 and enter any time extension(s) necessary to prevent this case from being abandoned.

Respectfully submitted,

Mark E. Scott
PTO Reg. No. 43,100
CONLEY ROSE, P.C.
(713) 238-8000 (Phone)
(713) 238-8008 (Fax)

ATTORNEY FOR APPLICANTS